View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by VBN

Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Profile likelihood in directed graphical models from BUGS output

Højbjerre, Malene

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Højbjerre, M. (2007). Profile likelihood in directed graphical models from BUGS output. Poster session presented at The Eighth International Workshop on Artificial Intelligence and Statistics, Key West, Florida, United States.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
? You may not further distribute the material or use it for any profit-making activity or commercial gain
? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: April 30, 2017

Profile Likelihood in Directed Graphical Models from BUGS Output

Malene Højbjerre, Department of Mathematical Sciences, Aalborg University, Denmark, *malene@math.auc.dk*

Introduction

Approximate profile likelihood of parameters/functions of parameters in directed graphical models with incomplete data from posterior samples from software BUGS (Bayesian inference Using Gibbs Sampling).

Directed Graphical Model

Directed graphical model defined by

- directed acyclic graph, $\mathcal{G} = (V, E)$
- joint probability distribution of $\boldsymbol{v} = (v_v)_{v \in V}$ that is directed Markov w.r.t. to $\mathcal{G}$

i.e. admit *recursive factorization property*

$$p(\boldsymbol{v}) = \prod_{v \in V} p(v_v | \boldsymbol{v}_{\mathrm{pa}(v)})$$

Assume

$$V = X \cup Y \cup \Theta \cup C$$

	random (single-edged)	constant (double-edged)
observed (rectangle)	X: observed data, $\boldsymbol{x} = (x_v)_{v \in X}$	C: covariates, $\boldsymbol{c} = (c_v)_{v \in C}$
unobserved (circle)	Y: missing data/ latent variables, $\boldsymbol{y} = (y_v)_{v \in Y}$	Θ: parameters, $\boldsymbol{\theta} = (\theta_v)_{v \in \Theta}$

As Spiegelhalter (1998) consider constants as random variables with priors on, but condition on them.

Let

$$z_v = \begin{cases} x_v & \text{for } v \in X \\ y_v & \text{for } v \in Y \end{cases}$$

$$\boldsymbol{z}_{\mathrm{pa}(v)} = (\boldsymbol{x}_{\mathrm{pa}(v)}, \boldsymbol{y}_{\mathrm{pa}(v)})$$

Then

$$p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{c}) = \prod_{v \in X \cup Y} p(z_v | \boldsymbol{z}_{\mathrm{pa}(v)}, \boldsymbol{\theta}_{\mathrm{pa}(v)}, \boldsymbol{c}_{\mathrm{pa}(v)}) \times \prod_{v \in \Theta} p(\theta_v) \prod_{v \in C} p(c_v)$$

Since $\boldsymbol{\theta}$ and $\boldsymbol{c}$ mutually independent

$$p(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{\theta}, \boldsymbol{c}) = \prod_{v \in X \cup Y} p(z_v | \boldsymbol{z}_{\mathrm{pa}(v)}, \boldsymbol{\theta}_{\mathrm{pa}(v)}, \boldsymbol{c}_{\mathrm{pa}(v)})$$

Likelihood of $\boldsymbol{\theta}$

$$L(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c}) = \int p(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{\theta}, \boldsymbol{c}) d\boldsymbol{y}$$

Profile likelihood of θ_i, $i \in \Theta$,

$$\hat{L}(\theta_i | \boldsymbol{x}, \boldsymbol{c}) = \sup_{\boldsymbol{\theta}_{\setminus i}} L(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c})$$

where $\boldsymbol{\theta}_{\setminus i} = \boldsymbol{\theta}_{\Theta \setminus \{i\}}$.

Profile likelihood of $\phi = g(\boldsymbol{\theta})$

$$\hat{L}(\phi | \boldsymbol{x}, \boldsymbol{c}) = \sup_{\boldsymbol{\theta} \in g^{-1}(\phi)} L(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c}), \quad \phi \in \Phi$$

where Φ image of parameter space of $\boldsymbol{\theta}$ under g.

Integrated likelihood of θ_i, $i \in \Theta$,

$$\begin{aligned} \check{L}(\theta_i | \boldsymbol{x}, \boldsymbol{c}) &= \int L(\theta_i, \boldsymbol{\theta}_{\setminus i} | \boldsymbol{x}, \boldsymbol{c}) p(\boldsymbol{\theta}_{\setminus i} | \theta_i) d\boldsymbol{\theta}_{\setminus i} \\ &= \int p(\boldsymbol{x}, \boldsymbol{w} | \theta_i, \boldsymbol{c}) d\boldsymbol{w} \end{aligned}$$

where $\boldsymbol{w} = (\boldsymbol{y}, \boldsymbol{\theta}_{\setminus i})$.

Method

Version 1

$\boldsymbol{\theta}_0$: fixed value of $\boldsymbol{\theta}$. As Geyer & Thompson (1992) consider

$$\frac{p(\boldsymbol{x} | \boldsymbol{\theta}, \boldsymbol{c})}{p(\boldsymbol{x} | \boldsymbol{\theta}_0, \boldsymbol{c})} = \frac{1}{\alpha} L(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c})$$

Rewrite likelihood function

$$L(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c}) \propto \int \prod_{v \in \mathrm{ch}(\Theta)} \frac{p(z_v | \boldsymbol{z}_{\mathrm{pa}(v)}, \boldsymbol{\theta}_{\mathrm{pa}(v)}, \boldsymbol{c}_{\mathrm{pa}(v)})}{p(z_v | \boldsymbol{z}_{\mathrm{pa}(v)}, \boldsymbol{\theta}_{0\,\mathrm{pa}(v)}, \boldsymbol{c}_{\mathrm{pa}(v)})} p(\boldsymbol{y} | \boldsymbol{x}, \boldsymbol{\theta}_0, \boldsymbol{c}) d\boldsymbol{y}$$

Draw $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(N)}$ from $p(\boldsymbol{y} | \boldsymbol{x}, \boldsymbol{\theta}_0, \boldsymbol{c})$.

Approximate likelihood function

$$\tilde{L}(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c}) \propto \sum_{j=1}^{N} \prod_{v \in \mathrm{ch}(\Theta)} \frac{p(z_v^{(j)} | \boldsymbol{z}_{\mathrm{pa}(v)}^{(j)}, \boldsymbol{\theta}_{\mathrm{pa}(v)}, \boldsymbol{c}_{\mathrm{pa}(v)})}{p(z_v^{(j)} | \boldsymbol{z}_{\mathrm{pa}(v)}^{(j)}, \boldsymbol{\theta}_{0\,\mathrm{pa}(v)}, \boldsymbol{c}_{\mathrm{pa}(v)})}$$

where $z_v^{(j)} = x_v$, for $v \in X$, $z_v^{(j)} = y_v^{(j)}$, for $v \in Y$, and $\boldsymbol{z}_{\mathrm{pa}(v)}^{(j)} = (\boldsymbol{x}_{\mathrm{pa}(v)}, \boldsymbol{y}_{\mathrm{pa}(v)}^{(j)})$.

Sampling from $p(\boldsymbol{y} | \boldsymbol{x}, \boldsymbol{\theta}_0, \boldsymbol{c})$: in BUGS with $\boldsymbol{\theta}$ fixed as $\boldsymbol{\theta}_0$.

$\boldsymbol{\theta}_0$: posterior mean of initial Gibbs sampler with priors on $\boldsymbol{\theta}$.

Compute $\log \tilde{L}(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c})$ in grid formed by quantiles of initial Gibbs sampler.

Maximize over grid to approximate profile log-likelihood.

Version 2

$\boldsymbol{\psi}$: another generic symbol of $\boldsymbol{\theta}$. Rewrite likelihood function

$$L(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c}) \propto \iint \prod_{v \in \mathrm{ch}(\Theta)} \frac{p(z_v | \boldsymbol{z}_{\mathrm{pa}(v)}, \boldsymbol{\theta}_{\mathrm{pa}(v)}, \boldsymbol{c}_{\mathrm{pa}(v)})}{p(z_v | \boldsymbol{z}_{\mathrm{pa}(v)}, \boldsymbol{\psi}_{\mathrm{pa}(v)}, \boldsymbol{c}_{\mathrm{pa}(v)})} p(\boldsymbol{y}, \boldsymbol{\psi} | \boldsymbol{x}, \boldsymbol{c}) d\boldsymbol{y} d\boldsymbol{\psi}$$

Draw $(\boldsymbol{y}^{(1)}, \boldsymbol{\psi}^{(1)}), (\boldsymbol{y}^{(2)}, \boldsymbol{\psi}^{(2)}), \ldots, (\boldsymbol{y}^{(N)}, \boldsymbol{\psi}^{(N)})$ from $p(\boldsymbol{y}, \boldsymbol{\psi} | \boldsymbol{x}, \boldsymbol{c})$.

Approximate likelihood function

$$\tilde{\tilde{L}}(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c}) \propto \sum_{j=1}^{N} \prod_{v \in \mathrm{ch}(\Theta)} \frac{p(z_v^{(j)} | \boldsymbol{z}_{\mathrm{pa}(v)}^{(j)}, \boldsymbol{\theta}_{\mathrm{pa}(v)}, \boldsymbol{c}_{\mathrm{pa}(v)})}{p(z_v^{(j)} | \boldsymbol{z}_{\mathrm{pa}(v)}^{(j)}, \boldsymbol{\psi}_{\mathrm{pa}(v)}^{(j)}, \boldsymbol{c}_{\mathrm{pa}(v)})}$$

Sampling from $p(\boldsymbol{y}, \boldsymbol{\psi} | \boldsymbol{x}, \boldsymbol{c})$: in BUGS with priors on $\boldsymbol{\theta}$.

Compute $\log \tilde{\tilde{L}}(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c})$ in grid formed by quantiles of Gibbs sampler in this version.

Maximize over grid to approximate profile log-likelihood.

Profile likelihood of a function

$\tilde{\nabla}$: grid points already considered. Form pairs

$$(g(\boldsymbol{\theta}), \tilde{L}(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c})), \quad \boldsymbol{\theta} \in \tilde{\nabla}$$

Need to calculate $g(\boldsymbol{\theta})$, already have $\tilde{L}(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c})$.

Order pairs in increasing order w.r.t. $\phi = g(\boldsymbol{\theta})$, partition into bins.

Find pair with maximum value of $\tilde{L}(\boldsymbol{\theta} | \boldsymbol{x}, \boldsymbol{c})$ for each bin.

Combine maximum pairs to approximate profile log-likelihood of $g(\boldsymbol{\theta})$.

Integrated likelihood of θ_i

Consider $\Theta \setminus \{i\}$ as latent variables in method.

Get a 1-dimensional grid. Marginalisation done by summation instead of maximalisation.

Comments

Priors: no significant impact, once Gibbs sampler converged.

Method: hybrid of Bayesian and likelihood - prior used to compute credible region, approximation done there.

Complementary tool to BUGS.

Example

From Spiegelhalter, Thomas, Best & Gilks (1996b). Measure for ten power plant pumps

c_i : operation time

x_i : number of failures

Model: $i = 1, 2, \ldots, 10$

$$x_i | y_i, c_i \sim \mathrm{Po}(y_i c_i)$$

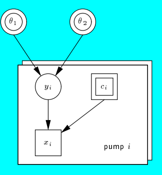
$$y_i | \theta_1, \theta_2 \sim \Gamma(\theta_1, \theta_2)$$

where

y_i : failure rate of pump i

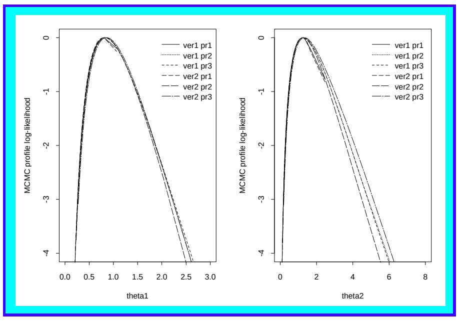
θ_1 : shape parameter

θ_2 : scale parameter

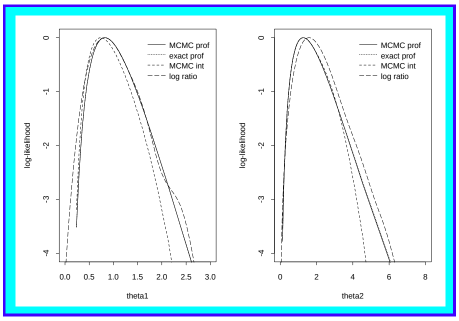


PRIORS	Prior 1	Prior 2	Prior 3
θ_1	$\mathrm{Exp}(10)$	$\mathrm{Exp}(1.0)$	$\mathrm{Exp}(0.01)$
θ_2	$\Gamma(0.1, 1.0)$	$\Gamma(0.01, 0.1)$	$\Gamma(0.001, 0.001)$

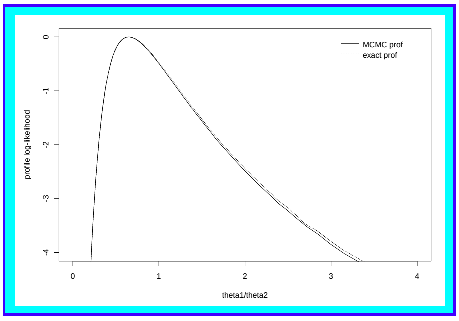
Sample size: $N = 5.000$ Grid size: 200×200



Priors no significant impact



Good approximation



Good approximation

References

Geyer, C. J. & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data, *Journal of the Royal Statistical Society, Series B* **54**(3): 657–699.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N. & Leimer, H.-G. (1990). Independence properties of directed Markov fields, *NETWORKS* **20**: 491–505.

Spiegelhalter, D. J. (1998). Bayesian graphical modelling: a case-study in monitoring health outcomes, *Applied Statistics* **47**: 115–133.

Spiegelhalter, D. J., Thomas, A., Best, N. G. & Gilks, W. (1996a). *BUGS 0.5 Bayesian inference Using Gibbs sampling Manual (version ii)*, MRC Biostatistics Unit, Cambridge.

Spiegelhalter, D. J., Thomas, A., Best, N. G. & Gilks, W. (1996b). *BUGS 0.5 Examples Volume 1 (version i)*, MRC Biostatistics Unit, Cambridge.